

VirHostNet 2.0: surfing on the web of virus/host molecular interactions data

Thibaut Guirimand¹, Stéphane Delmotte^{1,2} and Vincent Navratil^{1,*}

¹PRABI, Rhône Alpes Bioinformatics Center, UCBL, Lyon1, Université de Lyon, Lyon, France and ²UMR5558, UCBL, Lyon1, Université de Lyon, Lyon, France

Received September 22, 2014; Revised October 24, 2014; Accepted October 24, 2014

ABSTRACT

VirHostNet release 2.0 (<http://virhostnet.prabi.fr>) is a knowledgebase dedicated to the network-based exploration of virus–host protein–protein interactions. Since the previous VirHostNet release (2009), a second run of manual curation was performed to annotate the new torrent of high-throughput protein–protein interactions data from the literature. This resource is shared publicly, in PSI-MI TAB 2.5 format, using a PSICQUIC web service. The new interface of VirHostNet 2.0 is based on Cytoscape web library and provides a user-friendly access to the most complete and accurate resource of virus–virus and virus–host protein–protein interactions as well as their projection onto their corresponding host cell protein interaction networks. We hope that the VirHostNet 2.0 system will facilitate systems biology and gene-centered analysis of infectious diseases and will help to identify new molecular targets for antiviral drugs design. This resource will also continue to help worldwide scientists to improve our knowledge on molecular mechanisms involved in the antiviral response mediated by the cell and in the viral strategies selected by viruses to hijack the host immune system.

INTRODUCTION

Infectious diseases caused by viruses are responsible for millions of death every year as surveyed by the World Health Organization (<http://www.who.int>). This is the case of two emerging disease outbreaks caused by the pandemic Avian Influenza virus (H5N1) and more recently by the Ebola virus in 2014. The prediction, the prevention, the treatment and the understanding of such infectious diseases are today a priority for researchers.

Viruses are obligate intracellular pathogens that rely on the host cell machinery to achieve their own multiplication. Virus-encoded molecules tightly control the life cycle of viruses. It includes viral proteins that do not act alone during the infection but dynamically interact with different

cellular protein partners to express their functions. Nowadays, thousand viral genomes are fully sequenced and more than 2.5 million proteins are annotated in databases. However, their molecular and biological functions remain largely unknown or underdetermined. Since the beginning of the molecular biology revolution, numerous small-scale experiments were designed to better characterize the function mediated at the virus/host (VH) protein–protein interaction (ppi) interface. This trend has been emphasized in the last 10 years by the deluge of yeast two-hybrid (Y2H) and tandem affinity purification (TAP) data produced to characterize VH ppi in a systems biology perspective (1–6).

The VirHostNet knowledgebase release 1.0 (7) was launched in this context in 2007. The VirHostNet manual annotations were used during these last years as a gold standard in many VH interactome experiments, but also as a starting point for systems biology studies (2–3,8–13). Altogether, this resource has participated to improve the reproducibility of research in the field of VH interactomics, allowing scientists to compare their own experimental dataset and the properties observed from the underlying reconstructed protein interaction networks to those published previously. Hence, the VirHostNet resource was cited more than 70 times since its publication in the 2009 NAR database issue (source Google Scholar, September 2014). We have also demonstrated the scientific interest and the quality of such data, by providing new clues on convergent strategies selected by viruses to hijack and evade from the host innate immune functions (autophagy, response to type I interferon) as well as the role of viruses in the development of complex infectious diseases (2,8–9,12–13). Despite the growing numbers of virus/virus (VV) and VH ppi dataset published in the last 5 years (14), only few other large-scale annotation efforts were reported so far, excepted those performed by Intact (15), VirusMint (16,17) and a new resource, HPIDB (18).

In this issue, we present our last VirHostNet 2.0 curation update that leads to a 5-fold enrichment of the VH ppi data available in VirHostNet 1.0. Comparison to other resources shows that more than 80% of our annotations are unique to VirHostNet 2.0, emphasizing the high sensitivity of our curation approach. An implementation of a PSIC-

*To whom correspondence should be addressed. Tel: +33 472432826; Fax: +33 472431388; Email: navratil@prabi.fr

QUIC web server (19) is provided to make this dataset available in a standard way. In the VirHostNet 2.0, the model and the data were extended to non-human hosts and their associated viruses. These data were complemented with annotations extracted from the eight external databases using the PSICQUIC EBI web service. The biological relevance of our combined VH ppi dataset was evaluated by using VH species relationships annotations provided by UniProt (20) and Viralzone (21). Finally, we have also dramatically improved visual access to our unique resource, by developing a new web interface based on HMTL 5 technologies. This new development relies on a system-level modeling framework that significantly increases the interactivity between graph and tabular representation of VH ppi.

Biocuration and integration of VH ppi data

The VirHostNet 2.0 ppi data (September 2014) were obtained from multiple and complementary sources. First, intraspecies (VV and HH) and interspecies (VH) ppi were manually annotated from the scientific literature using the PSI-MI TAB 2.5 format (22). Hence, each ppi is manually annotated according to the spoke model, by using the accession numbers of the two protein partners (UniProt or Refseq primary accession number are used), the PSI-MI method accession number describing the experiment and the PubMed publication id (PMID). This approach has the advantage to rapidly increase the sensitivity of the VH and VV ppi dataset as compared to more accurate annotation strategies (PSI-MI TAB 2.6 and 2.7), without losing any information concerning the redundancy of experiments. Moreover, this solution will provide a solid foundation for the next runs of annotation in the richer PSI-MI TAB 2.6 or 2.7 formats. All the VirHostNet annotations are made publicly available at EBI, throughout a PSICQUIC web service (19). These data were next completed (update will follow the UniProt release frequency) with ppi data extracted from eight high-confidence external databases (including IntAct, Mint, DIP, InnateDB, BIND, UniProt, HPIDb) by using the EBI PSICQUIC web service. To overcome the heterogeneity of gene/protein accession numbers, all protein entries are first mapped onto UniProt (or NCBI/Refseq if not available) primary accession numbers by using the UniProt and NCBI cross-references. In the VirHostNet release 2.0, we also take into account viral protein domain interactions by using domain definition provided by UniProt and RefSeq (mature peptide). Finally, this combined ppi dataset was enriched with biological metadata, as well as protein functional annotations (NCBI Taxonomy, Gene Ontology, Kegg pathways, protein domain annotation) extracted from the UniProt flat files.

An up-to-date VH ppi dataset extended to non-human hosts

Our new round of biocuration results in a 5-fold enrichment according to the previous release 1.0 of VirHostNet. Over 30 000 ppi are now annotated, with a total of 16 000 VH, 2 500 VV and 12 000 host/host (HH) ppi (Figure 1a, Supplementary Table S1). A comparison to the data stored in known available databases reveals that more than 80% of annotated interactions are unique to the VirHostNet 2.0 resource, confirming the relatively high sensitivity of our fast

annotation approach. It is interesting to note that 50 and 42% of VH ppi data were obtained by TAP and Y2H experiment, respectively (Figure 1b). Moreover, more than 33% of the VirHostNet 2.0 ppi are supported by at least two independent experiments, suggesting the high quality of those interactions (Figure 1c).

The VH ppi were next flagged according to the information of VH species relationships provided by UniProt (20) and Viralzone (21). This annotation step allows us to better control the true nature of annotated interspecies interactions and to clearly differentiate ppi related to biological VH relationships from more artificial ones. Hence, we show that more than 90% of the VirHostNet 2.0 curated ppi corresponds to biological VH species relationships, suggesting the high specificity of the available data and annotations. Nevertheless, even when considering the combined dataset, we note a poor coverage of the VH interactions space (less than 2% of the 5958 VH relationships). Hence, at the viral 'species' rank (as defined by the NCBI taxonomy classification), only 107 known VH relationships are described by at least one molecular interaction. This corresponds to 259 viral taxons when considering the diversity of viral strains for each viral species. Moreover, at least one molecular interaction experiment has been described for 11 distinct host species (Supplementary Table S1), including five vertebrates, one invertebrate, three plants and one prokaryote, with a strong inspection bias toward virus/human ppi (99% of the ppi dataset). This result clearly reflects the research and the annotation efforts to identify the molecular basis of viral replication and pathogenesis in human. However, the number of non-yet annotated VH ppi is growing for non-human hosts (23), such as animals (*Bos Taurus*, *Sus scrofa*, *Gallus gallus*) and plants (*Arabidopsis thaliana*) of agronomical interest. Hence, to better anticipate this trend, we decided to extend the previous model of VirHostNet to all known viral hosts.

Toward system-level measurement in a VH protein interaction network framework

As described previously (8,9), the combined VH, VV and HH ppi are modeled by using undirected multicolored graph formalism, in order to complement our data with systems biology knowledge. More precisely, three kinds of graph are defined: the VH pin (VH—the set of ppi between viral and host proteins), the host pin (HH—the set of ppi between host proteins) and the VH pin projection onto the host pin (V2H—the set of ppi between viral and connected host proteins). We have previously demonstrated that host proteins interacting with viral proteins exhibit particular graph properties that might be correlated to emerging biological functions essential for the viral life cycle and involved in the physio-pathogenesis of viruses (8,9). Hence, a set of graph metrics (degree, betweenness, closeness, transitivity of nodes), frequently used in the fields of interactomics to better characterize network properties, were pre-computed with the igraph R package for each network (VH, HH, V2H pin) and stored into our knowledgebase. The graph visualization rendering and coloring scheme presented below is based on these three network models. Hence, visualization of the nodes and edges graph properties will

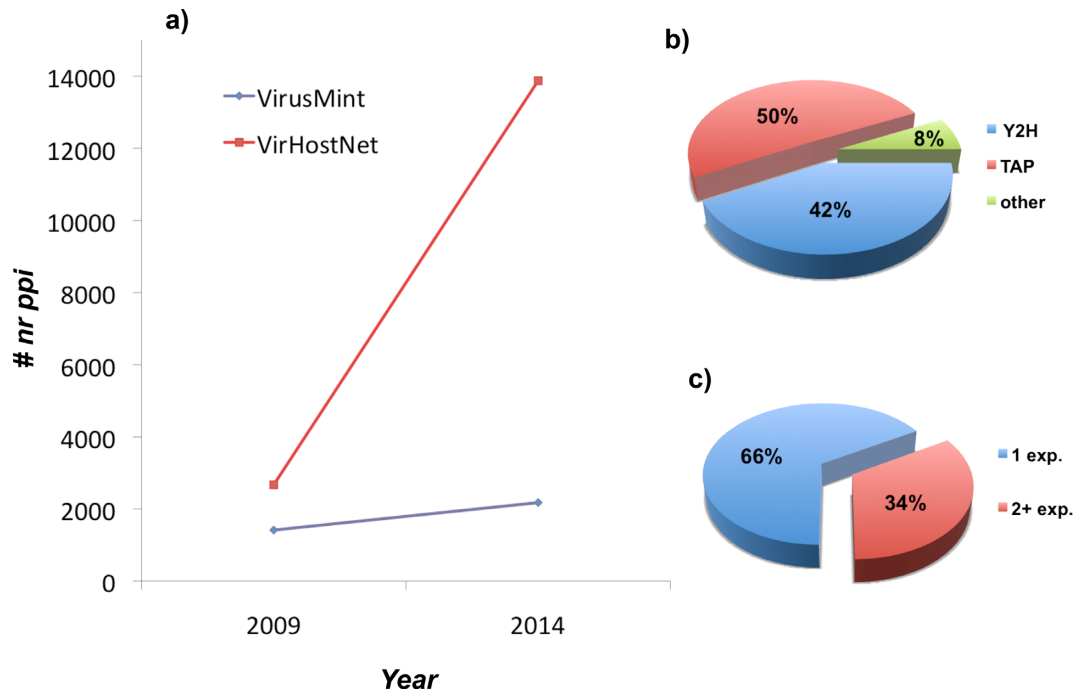


Figure 1. VirHostNet 2.0 annotations statistics. (a) Growth statistics. The number of non-redundant (i.e. unique) VH ppi annotated between 2009 and 2014 is given for VirHostNet and VirusMint. This figure highlights a 5-fold enrichment of the VirHostNet 2.0 annotation between 2009 and 2014. (b) Experiments summary. The pie chart represents the proportion of annotated experimental methods used to characterize VH and VV ppi (other—other methods). (c) Independent experiments support. This pie chart represents the proportion of VH and VV non-redundant ppi supported by one or more than two independent experiments.

help the users to rapidly highlight viral and host proteins as well as biologically relevant VH (i.e. those validated by VH species relationships), VV and HH ppi from large networks.

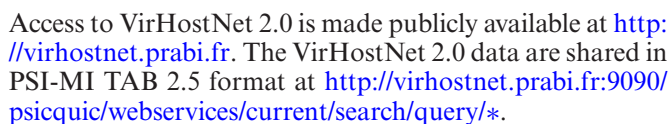
A brand new VH ppi network viewer

We have developed a new web-based VH ppi viewer (aka VirHostScope) based on Cytoscape web (24), jQuery and Bootstrap libraries. This web interface is available at <http://virhostnet.prabi.fr> and allows users to rapidly query and visualize proteins and ppi data in the context of VH, HH and V2H pin. A wiki documentation and a short tutorial are provided on the website.

The VirHostScope 'home' page is composed of (i) a 'quick-search' text area, (ii) a BLASTP (25) form and (iii) a taxonomic browser (see Figure 2 and the wiki for more details). The 'quick search' form allows—in 'simple search' mode—to retrieve direct partners of a protein query based on UniProt and RefSeq accession numbers. When selecting 'multiple search', the user can enter multiple protein accession numbers and visualize interactions among them. The web interface provides also functionalities to visualize protein interactions between a list of proteins annotated (i) at a NCBI taxonomic rank, (ii) with a specific domain composition (Interpro), (iii) acting in a specific pathway (KEGG) or (iv) annotated from a publication referenced by its PMID. All search forms are tuned with autocomplete functionality. In addition, a 'BLAST' form allows the user to make homology search against a database containing all the sequences of the protein partners stored in VirHostNet 2.0. This is a simple way to rapidly predict putative partners of

an unknown protein sequence or a variant, based on the assumption of interology (i.e. the conservation of ppi between two protein partners and their homologs) (see the wiki page for more details and a case study). The 'browser' page summarizes the number of ppi and partners available at each viral NCBI taxonomic rank (Baltimore group, Family rank, Species rank), and allows retrieving the corresponding protein interaction network.

In the 'visualization' page (Figure 2), we have developed highly efficient tools to reconstruct and to explore fully personalized protein interaction networks. We provide specific buttons at the top of the Cytoscape web page to explore viral and/or host partners of proteins or to connect a set of selected proteins. The 'add' option of the quick-search panel allows to increment recursively the current protein interaction network, by using protein, domains, pathways, taxonomy or PMID accessions. A 'table panel' enumerates and colors dynamically the list of proteins and ppi according to the current graph or its sub-selection. It also provides a useful taxonomy summary of the protein interaction network (the number of viral and host proteins, the number of VH, VV, HH ppi). We also propose a rich 'option panel' that helps users to colorize, resize and filter graph nodes (proteins) and edges (ppi) according to multiple and complex criteria, including (i) taxonomy, (ii) molecular interaction experiments (number of independent experiments, independent experimental methods, publications or independent database supporting the interaction) and (iii) graph properties of nodes and edges pre-computed in the three graph models (see section above). The pin generated by Cy-



SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was performed using the computing facilities of the CC LBBE/PRABI. We want to thank the EBI PSICQUIC team for registering our PSICQUIC web service, but also Marie Primout for her contribution to the biocuration and Sidoine, Justine, Charli and Marius Navratil for their critical reading of the manuscript.

FUNDING

The Fondation INNOVations en Infectiologie (FINOVI) [R12128CC]. Funding for open access charge: FINOVI [R12128CC].

Conflict of interest statement. None declared.

REFERENCES

- Uetz,P., Dong,Y.A., Zeretzke,C., Atzler,C., Baiker,A., Berger,B., Rajagopala,S.V., Roupelieva,M., Rose,D., Fossum,E. *et al.* (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, **311**, 239–242.
- de Chassey,B., Navratil,V., Tafforeau,L., Hiet,M.S., Aublin-Gex,A., Agaue,S., Meiffren,G., Pradezynski,F., Faria,B.F., Chantier,T. *et al.* (2008) Hepatitis C virus infection protein network. *Mol. Syst. Biol.*, **4**, 230.
- Griffiths,S.J., Koegl,M., Boutell,C., Zenner,H.L., Crump,C.M., Pica,F., Gonzalez,O., Friedel,C.C., Barry,G., Martin,K. *et al.* (2013) A systematic analysis of host factors reveals a Med23-interferon-lambda regulatory axis against herpes simplex virus type 1 replication. *PLoS Pathog.*, **9**, e1003514.
- Gulbahce,N., Yan,H., Dricot,A., Padi,M., Byrdsong,D., Franchi,R., Lee,D.S., Rozenblatt-Rosen,O., Mar,J.C., Calderwood,M.A. *et al.* (2012) Viral perturbations of host networks reflect disease etiology. *PLoS Comput. Biol.*, **8**, e1002531.
- Calderwood,M.A., Venkatesan,K., Xing,L., Chase,M.R., Vazquez,A., Holthaus,A.M., Ewence,A.E., Li,N., Hirozane-Kishikawa,T., Hill,D.E. *et al.* (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 7606–7611.
- König,R., Stertz,S., Zhou,Y., Inoue,A., Hoffmann,H.H., Bhattacharyya,S., Alamares,J.G., Tscherne,D.M., Ortigoza,M.B., Liang,Y. *et al.* (2010) Human host factors required for influenza virus replication. *Nature*, **463**, 813–817.
- Navratil,V., de Chassey,B., Meyniel,L., Delmotte,S., Gautier,C., Andre,P., Lotteau,V. and Rabourdin-Combe,C. (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.*, **37**, D661–D668.
- Navratil,V., de Chassey,B., Combe,C.R. and Lotteau,V. (2011) When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst. Biol.*, **5**, 13.
- Navratil,V., de Chassey,B., Meyniel,L., Pradezynski,F., Andre,P., Rabourdin-Combe,C. and Lotteau,V. (2010) System-level comparison of protein-protein interactions between viruses and the human type I interferon system network. *J. Proteome Res.*, **9**, 3527–3536.
- de Chassey,B., Aublin-Gex,A., Ruggieri,A., Meyniel-Schicklin,L., Pradezynski,F., Davoust,N., Chantier,T., Tafforeau,L., Mangeot,P.E., Ciancia,C. *et al.* (2013) The interactomes of influenza virus NS1 and NS2 proteins identify new host factors and provide insights for ADAR1 playing a supportive role in virus replication. *PLoS Pathog.*, **9**, e1003440.
- Simonis,N., Rual,J.F., Lemmens,I., Boxus,M., Hirozane-Kishikawa,T., Gatot,J.S., Dricot,A., Hao,T., Vertommen,D., Legros,S. *et al.* (2012) Host-pathogen interactome mapping for HTLV-1 and -2 retroviruses. *Retrovirology*, **9**, 26.
- de Chassey,B., Meyniel-Schicklin,L., Aublin-Gex,A., Navratil,V., Chantier,T., Andre,P. and Lotteau,V. (2013) Structure homology and interaction redundancy for discovering virus-host protein interactions. *EMBO Rep.*, **14**, 938–944.
- Gregoire,I.P., Richetta,C., Meyniel-Schicklin,L., Borel,S., Pradezynski,F., Diaz,O., Deloire,A., Azocar,O., Baguet,J., Le Breton,M. *et al.* (2011) IRGM is a common target of RNA viruses that subvert the autophagy network. *PLoS Pathog.*, **7**, e1002422.
- Rozenblatt-Rosen,O., Deo,R.C., Padi,M., Adelmant,G., Calderwood,M.A., Rolland,T., Grace,M., Dricot,A., Askenazi,M., Tavares,M. *et al.* (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature*, **487**, 491–495.
- Kerrien,S., Aranda,B., Breuza,L., Bridge,A., Broackes-Carter,F., Chen,C., Duesbury,M., Dumousseau,M., Feuermann,M., Hinz,U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, **40**, D841–D846.
- Chatr-aryamontri,A., Ceol,A., Peluso,D., Nardozza,A., Panni,S., Sacco,F., Tinti,M., Smolyar,A., Castagnoli,L., Vidal,M. *et al.* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res.*, **37**, D669–D673.
- Calderone,A., Licata,L. and Cesareni,G. (2014) VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res.*, pii:gku830.
- Kumar,R. and Nanduri,B. (2010) HPIDB—a unified resource for host-pathogen interactions. *BMC Bioinformatics*, **11**, S16.
- del-Toro,N., Dumousseau,M., Orchard,S., Jimenez,R.C., Galeota,E., Launay,G., Goll,J., Breuer,K., Ono,K., Salwinski,L. *et al.* (2013) A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.*, **41**, W601–W606.
- The Uniprot Consortium. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
- Masson,P., Hulo,C., De Castro,E., Bitter,H., Gruenbaum,L., Essioux,L., Bougueleret,L., Xenarios,I. and Le Mercier,P. (2013) ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Res.*, **41**, D579–D583.
- Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A.F., Vinod,N., Bader,G.D., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.
- Wessling,R., Epple,P., Altmann,S., He,Y., Yang,L., Henz,S.R., McDonald,N., Wiley,K., Bader,K.C., Glasser,C. *et al.* (2014) Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe*, **16**, 364–375.
- Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.